The Annals of Applied Statistics
2012, Vol. 6, No. 4, 1814–1837
DOI: 10.1214/12-AOAS555

ADDRESSING MISSING DATA MECHANISM UNCERTAINTY USING MULTIPLE-MODEL MULTIPLE IMPUTATION: APPLICATION TO A LONGITUDINAL CLINICAL TRIAL

BY JUNED SIDDIQUE[1], OFER HAREL[2] AND CATHERINE M. CRESPI[3]

Northwestern University, University of Connecticut and University of California Los Angeles

We present a framework for generating multiple imputations for continuous data when the missing data mechanism is unknown. Imputations are generated from more than one imputation model in order to incorporate uncertainty regarding the missing data mechanism. Parameter estimates based on the different imputation models are combined using rules for nested multiple imputation. Through the use of simulation, we investigate the impact of missing data mechanism uncertainty on post-imputation inferences and show that incorporating this uncertainty can increase the coverage of parameter estimates. We apply our method to a longitudinal clinical trial of low-income women with depression where nonignorably missing data were a concern. We show that different assumptions regarding the missing data mechanism can have a substantial impact on inferences. Our method provides a simple approach for formalizing subjective notions regarding nonresponse so that they can be easily stated, communicated and compared.

1. Introduction. The longitudinal clinical trial is a powerful design for estimating and comparing rates of change over time in two or more treatment groups. However, measuring participants repeatedly over time provides repeated opportunities for participants to miss measurement occasions. Missing values are a problem in most longitudinal studies and a variety of methods have been developed to produce valid inferences in the presence of missing data. In particular, multiple imputation—where missing values are

Received April 2011; revised January 2012.
[1]Supported in part by AHRQ Grant R03 HS018815 and NIH Grants R01 MH040859 and K07 CA154862.
[2]Supported in part by NIH Grant K01 MH087219.
[3]Supported in part by NIH Grants CA133081 and UL1TR000124.
Key words and phrases. Nonignorable, NMAR, MNAR, not missing at random, missing not at random.

replaced with two or more plausible values—has gained widespread acceptance in recent years and is a common and flexible approach for handling missing data.

When dealing with missing data, special concern must be given to the process that gave rise to the missing data, referred to as the missing data mechanism. Most methods for generating multiple imputations, both fully-parametric methods [Liu (1995), Schafer (1997)] and semi-parametric methods [Raghunathan et al. (2001), Schenker and Taylor (1996), Siddique and Belin (2008a), van Buuren (2007)], assume the missing data mechanism is *ignorable* as described by Rubin (1976), where the probability that a value is missing does not depend on unobserved information such as the value itself. When data are nonignorably missing, that is, the probability that a value is missing *does* depend on unobserved information, the model for generating imputations must take into account the missing data mechanism. The role of nonignorability assumptions has been discussed in the context of a variety of applied settings; see, for example, Little and Rubin [(2002), chapter 15], Belin et al. (1993), Rubin, Stern and Vehovar (1995), Schafer and Graham (2002), Wachter (1993) and Demirtas and Schafer (2003).

Nonignorably missing data is of particular concern in depression trials because it is very likely that the reason for a participant missing an assessment or dropping out of a study is related to their underlying depression status [Blackburn et al. (1981), Elkin et al. (1989), Warden et al. (2009)]. For example, a depressed participant may feel like the intervention is not working for them and may be unwilling to sit through an interview and/or answer the phone. Conversely, a high-functioning, nondepressed participant may feel like he no longer needs to remain in the trial or may not be available for an assessment because he is busy working, shopping or socializing. Failure to take into account the missing data mechanism may result in inferences that make a treatment appear more or less effective. Failure to incorporate uncertainty regarding the missing data mechanism may result in inferences that are overly precise given the amount of available information [Demirtas and Schafer (2003)].

Since a nonignorable missing data mechanism depends on unobserved data, there is little information available to correctly model this process. A common approach in such cases is to perform a sensitivity analysis, drawing inferences based on a variety of assumptions regarding the missing data mechanism [Daniels and Hogan (2008)]. There is a broad literature on sensitivity analyses for exploring unverifiable missing data assumptions [see Ibrahim and Molenberghs (2009) and discussion for a review]. One approach begins with the specification of a full-data distribution, followed by examination of inferences across a range of values for one or more unidentified parameters [Daniels and Hogan (2008), Molenberghs, Kenward and Goetghebeur (2001), Rubin (1977), Scharfstein, Rotnitzky and Robins (1999), Vansteelandt et al. (2006)].

When a decision is required, a drawback of sensitivity analysis is that it produces a range of answers rather than a single answer [Scharfstein, Rotnitzky and Robins (1999)]. Several authors have proposed model-based methods for obtaining a final inference. This approach involves placing an informative prior distribution on the unidentified parameters that characterize assumptions about the missing data mechanism. Then, inferences are drawn that incorporate a range of assumptions regarding the missing data mechanism [Daniels and Hogan (2008), Forster and Smith (1998), Kaciroti et al. (2006), Rubin (1977)].

An alternative approach for handling data with nonignorable missingness is multiple imputation. Multiple imputation methods have several advantages over model-based methods for analyzing data with missing values: they allow for standard complete-data methods of analysis to be performed once the data have been imputed [Little and Rubin (2002)], and auxiliary variables that are not part of the analysis procedure can be incorporated into the imputation procedure to increase efficiency and reduce bias [Collins, Schafer and Kam (2001)].

Methods for multiple imputation with nonignorably missing data include those of Carpenter, Kenward and White (2007) who use a reweighting approach to investigate the influence of departures from the ignorable assumption on parameter estimates. van Buuren, Boshuizen and Knook (1999) perform a sensitivity analysis with multiply imputed data using offsets to explore how robust their inferences are to violations of the assumption of ignorability. A limitation of these approaches is that they do not take into account uncertainty regarding the missing data mechanism. Instead, they provide a range of inferences for various ignorability assumptions.

Landrum and Becker (2001) develop an imputation procedure that allows for model uncertainty to be reflected in the multiple imputations for those cases in which no one imputation model is clearly the best model by drawing imputations from more than one model. However, their procedure assumes ignorably missing data. Siddique and Belin (2008b) use a nonignorable approximate Bayesian bootstrap to generate multiple imputations assuming nonignorability. Each set of imputations is based on a different assumption regarding the missing data mechanism in order to incorporate missing data mechanism uncertainty. However, Siddique and Belin (2008b) use conventional multiple imputation combining rules which are not appropriate when imputations are generated from different posterior distributions because they do not take into account the additional uncertainty due to using more than one imputation model.

In this paper we describe a new multiple imputation approach for estimating parameters and their associated confidence intervals in the presence of nonignorable nonresponse. Our goal is to develop a multiple imputation framework analogous to model-based methods such as those of Rubin (1977), Forster and Smith (1998) and Daniels and Hogan (2008) that incorporate a

range of ignorability assumptions into one inference. Rather than attempting the hopeless objective of correctly modeling the missing data mechanism, we generate our imputations using multiple imputation models and then use specialized combining rules to generate inferences that incorporate missing data mechanism uncertainty. Imputations are generated in three steps: (1) a distribution of models incorporating ignorable and/or nonignorable mechanisms is specified; (2) a model is drawn from this distribution; (3) multiple imputations are generated from the model selected in Step 2. Steps 2 and 3 are then repeated, thereby generating multiple-model multiple imputations. The nested imputation combining rules of Shen (2000) are used to combine inferences across multiple imputations so that between-model uncertainty is incorporated into the standard errors of parameter estimates.

The outline for the rest of this paper is as follows. In Section 2 we describe the WECare study, a longitudinal depression treatment trial that motivated this work. In Section 3 we describe methods for generating multiple-model multiple imputations for continuous data in order to incorporate missing data mechanism uncertainty and describe the nested imputation combining the rules of Shen (2000). In addition, we develop a method of quantifying the contribution of missing data mechanism uncertainty to the overall rate of missing information. Section 4 describes the design of a simulation study and Section 5 presents the results of the simulation study. In Section 6 we apply our approach to the WECare study. Section 7 provides a discussion.

Closely related to the concept of ignorability are the missing data mechanism taxonomies "missing at random" (MAR) and "not missing at random" (NMAR). MAR requires that the probability of missingness depends on observed values only, while ignorability includes the additional assumption that the parameters that generate the data and the parameters governing the missing data mechanism are distinct [Little and Rubin (2002), Rubin (1976)]. While distinctness of these two sets of parameters cannot always be assumed (particularly in time to event data), for the purposes of this paper we will use the terms MAR and ignorable interchangeably and the terms NMAR and nonignorable interchangeably.

2. Motivating example: The WECare study. The Women Entering Care (WECare) Study investigated depression outcomes during a 12-month period in which 267 low-income mostly minority women in the suburban Washington, DC area were treated for depression. The participants were randomly assigned to one of three treatment groups: Medication, Cognitive Behavioral Therapy (CBT) or treatment-as-usual (TAU), which consisted of referral to a community provider. Depression was measured every month through a phone interview using the Hamilton Depression Rating Scale (HDRS).

Information on ethnicity, income, number of children, insurance and education was collected during the screening and the baseline interviews. All

screening and baseline data were complete except for income, with 10 participants missing data on income. After baseline, the percentage of missing interviews ranged between 24% and 38% across months.

Outcomes for the first six months of the study were reported in Miranda et al. (2003). In that paper the primary research question was whether the Medication and CBT treatment groups had better depression outcomes compared to the TAU group. To answer this question, the data were analyzed on an intent-to-treat basis using a random intercept and slope regression model which controlled for ethnicity and baseline depression. Results from the complete-case analysis showed that both the Medication intervention ($p < 0.001$) and the CBT intervention ($p = 0.006$) reduced depression symptoms more than the TAU community referral.

This analysis assumed missing WECare values were MAR. An underlying concern was whether missing values were nonginorably missing. The motivation of the work described here was to develop methods of inference that would reflect uncertainty about the missing data mechanism in the WECare trial.

3. Methods. Our approach proceeds in four stages. First, a distribution of imputation models is specified. Then, nested imputation is conducted in which M models are drawn from this distribution of models and N multiple imputations for each missing value are generated from each of the M models resulting in $M \times N$ complete data sets. Next, parameters of interest are estimated along with their standard errors for each imputed data set. Finally, the parameter estimates and standard errors are combined using rules for nested multiple imputation that yield final inferential results. We also present a method of quantifying the contribution of missing data mechanism uncertainty to the overall rate of missing information.

3.1. *Specifying the distribution of imputation models.* The first step in our procedure is identifying a distribution of models from which it is possible to sample. The choice of which model to use will depend on subjective notions regarding the dissimilarity of observed and missing values that the imputer wishes to formalize. Ideally, this external information is elicited from experts or those who collected the data.

Rubin (1987) notes the importance of using easily communicated models to generate multiple imputations assuming nonignorability so that users of the completed data can make judgments regarding the relative merits of the various inferences reached under different nonresponse models. In this section we describe in detail a method for generating multiple imputations from multiple models using an adaptation of a nonignorable imputation procedure suggested by Rubin [(1987), page 22]. In the discussion section we discuss the application of our multiple model framework using other procedures.

3.2. *Transforming imputed ignorable continuous values to create nonignorable values.* Rubin [(1987), page 203] describes a simple transformation for generating nonignorable imputed values from ignorable imputed values:

$$\text{(nonignorable imputed } Y_i) = k \times \text{(ignorable imputed } Y_i). \tag{3.1}$$

For example, if $k = 1.2$, then the assumption is that, conditioning on other observed information, missing values are 20% larger than observed values. In order to create a distribution of nonignorable (and ignorable) models, we replace the multiplier k in equation (3.1) with multiple draws from some distribution. If the imputer believes that missing values tend to be larger than observed values, then a potential distribution for k might be a Uniform$(1, 3)$ distribution or a Normal$(1.5, 1)$ distribution. By centering the distribution of k around values smaller than 1.0, nonignorable imputations can be generated which assume that missing values are smaller than observed values after conditioning on observed information.

When the ignorable imputed value in equation (3.1) is negative, the right-hand side of the equation needs to be modified so that values of k greater than 1 will increase the value of the ignorable imputed value and values of k less than 1 will decrease the value of the ignorable imputed value. A more general version of equation (3.1), applicable in all settings, is

$$\begin{aligned}&\text{(nonignorable imputed } Y_i)\\ &\quad = [(k-1) \times |\text{ignorable imputed } Y_i|] + \text{ignorable imputed } Y_i.\end{aligned} \tag{3.2}$$

Caution should be exercised to avoid unrealistic imputations. Multipliers of large magnitude may result in imputations outside the range of plausible values.

If the imputer wants to generate imputations that are centered around a missing at random mechanism but with additional uncertainty, they could specify a Uniform$(0.5, 1.5)$ or Normal$(1.0, 0.25)$ distribution for the multiplier. More generally, Daniels and Hogan (2008) categorize the priors used in a sensitivity analysis as departures from a MAR mechanism. They use the following categories: MAR with no uncertainty, MAR with uncertainty, NMAR with no uncertainty and NMAR with uncertainty. When viewed in this framework, the standard MAR assumption (MAR with no uncertainty) is simply one mechanism across a continuum of mechanism specifications and is equivalent to using a Normal$(1, 0)$ or Uniform$(1, 1)$ distribution for the multiplier k in equation (3.2). Note that when we use the term "imputation model uncertainty" we are referring to uncertainty in the missing data mechanism as governed by uncertainty in the multiplier k.

When the data are continuous, equation (3.2) can be applied to ignorable imputed values that are generated from any imputation method that assumes ignorability. In this paper we generate ignorable imputations using

regression imputation [Rubin (1987), page 166]. We use different values for the multiplier k in equation (3.2) to easily generate imputations from many different models.

3.3. *Nested multiple imputation.* Once the distribution of models has been specified, imputation proceeds in two stages. First M models are drawn from a distribution of models such as those described in Section 3.2. Then N multiple imputations for each missing value are generated for each of the M models, resulting in $M \times N$ complete data sets.

More specifically, let the complete data be denoted by $Y = (Y_{\text{obs}}, Y_{\text{mis}})$. For the first stage, the imputation model ψ is drawn from its predictive distribution

$$\psi^m \sim p(\psi), \qquad m = 1, 2, \ldots, M. \tag{3.3}$$

The second stage starts with each model ψ^m and draws n independent imputations conditional on ψ^m,

$$Y_{\text{mis}}^{(m,n)} \sim p(Y_{\text{mis}}|Y_{\text{obs}}, \psi^m), \qquad n = 1, 2, \ldots, N. \tag{3.4}$$

Because the $M \times N$ nested multiple imputations are not independent draws from the same posterior predictive distribution of Y_{mis}, the traditional multiple imputation combining rules of Rubin (1987) do not apply. Instead, it is necessary to use combining rules that take into account variability due to the multiple models. Fortunately, the method described here is similar to nested multiple imputation [Harel (2007, 2009), Rubin (2003), Shen (2000)]. In the Appendix we provide further justification for using the nested imputation combining rules.

3.4. *Combining rules for final inference.* In this section we describe the nested multiple imputation combining rules that we use to combine inferences across multiply imputed data sets based on multiple imputation models. In describing the rules below, we use notation that follows closely to that of Shen (2000).

Let Q be the quantity of interest. Assume with complete data, inference about Q would be based on the large sample statement that

$$(Q - \hat{Q}) \sim N(0, U),$$

where $\hat{Q}$ is a complete-data statistic estimating Q and U is a complete-data statistic providing the variance of $Q - \hat{Q}$. The $M \times N$ imputations are used to construct $M \times N$ completed data sets, where the estimate and variance of Q from the single imputed data set is denoted by $(\hat{Q}^{(m,n)}, U^{(m,n)})$, where $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$. The superscript (m, n) represents the nth imputed data set under model m. Let $\bar{Q}$ be the overall average of all

$M \times N$ point estimates

$$\bar{Q} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{Q}^{(m,n)}, \tag{3.5}$$

and let $\bar{Q}_m$ be the average of the mth model,

$$\bar{Q}_m = \frac{1}{N} \sum_{n=1}^{N} \hat{Q}^{(m,n)}. \tag{3.6}$$

Three sources of variability contribute to the uncertainty in Q. These three sources of variability are as follows: $\bar{U}$, the overall average of the associated variance estimates

$$\bar{U} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} U^{(m,n)}, \tag{3.7}$$

W, the within-model variance

$$W = \frac{1}{M(N-1)} \sum_{m=1}^{M} \sum_{n=1}^{N} (\hat{Q}^{(m,n)} - \bar{Q}_m)^2, \tag{3.8}$$

and B, the between-model variance

$$B = \frac{1}{M-1} \sum_{m=1}^{M} (\bar{Q}_m - \bar{Q})^2. \tag{3.9}$$

The quantity

$$T = \bar{U} + \left(1 + \frac{1}{M}\right) B + \left(1 - \frac{1}{N}\right) W \tag{3.10}$$

estimates the total variance of $(Q - \bar{Q})$. Interval estimates and significance levels for scalar Q are based on a Student-t reference distribution

$$T^{-1/2}(Q - \bar{Q}) \sim t_v, \tag{3.11}$$

where v, the degrees of freedom, follows from

$$v^{-1} = \left[\frac{(1 + 1/M)B}{T}\right]^2 \frac{1}{M-1} + \left[\frac{(1 - 1/N)W}{T}\right]^2 \frac{1}{M(N-1)}. \tag{3.12}$$

In standard multiple imputation, only one model is used to generate imputations so that the between-model variance B [equation (3.9)] is equal to 0 and it is not necessary to account for the extra source of variability due to model uncertainty.

3.5. *Rates of missing information.* Standard multiple imputation provides a rate of missing information that may be used as a diagnostic measure of how the missing data contribute to the uncertainty about Q, the parameter of interest [Schafer (1997)]. Harel (2007, 2009) derived rates of

missing information for nested multiple imputation based on the amount of missing information due to model uncertainty and missingness. These rates include an overall rate of missing information γ, which can be partitioned into a between-model rate of missing information γ^b, and a within-model rate of missing information γ^w. With no missing information (either due to nonresponse or imputation model uncertainty), the variance of $(Q-\bar{Q})$ reduces to $\bar{U}$ so that the estimated overall rate of missing information is [Harel (2007)]

$$\hat{\gamma} = \frac{B + (1 - 1/N)W}{\bar{U} + B + (1 - 1/N)W}. \tag{3.13}$$

If the correct imputation model is known, then B, the between-model variance, is 0 and the estimated rate of missing information due to nonresponse is

$$\hat{\gamma}^w = \frac{W}{\bar{U} + W}. \tag{3.14}$$

Roughly speaking, equation (3.13) measures the fraction of total variance accounted for by nonresponse and model uncertainty and equation (3.14) measures the fraction of total variance accounted for by nonresponse when the correct imputation model is known. See Harel (2007, 2009) for details. The estimated rate of missing information due to model uncertainty is then $\hat{\gamma}^b = \hat{\gamma} - \hat{\gamma}^w$.

In a nested imputation framework, Harel (2008) takes the ratio $\frac{\hat{\gamma}^w}{\hat{\gamma}}$ which he terms *outfluence*. In nested imputation, outfluence is a measure of the influence of one type of missing data relative to all missing values. Here, we use the ratio $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ to measure the contribution of model uncertainty to the overall rate of missing information. For example, a value of $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ equal to 0.5 would suggest that half of the overall rate of missing information is due to missing data mechanism uncertainty, the other half due to missing values. We anticipate that most researchers would not want to exceed this value unless they have very little confidence in their imputation model. Note that most imputation procedures use one model and implicitly assume that $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ is equal to 0.

In the next section we present simulations showing that incorporating more than one imputation model in an imputation procedure increases both $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ and increases the coverage of parameter estimates versus procedures that use only one imputation model.

4. Design of simulation study. In this section we describe a simulation study to illustrate the method of multiple-model multiple imputation. We simulate longitudinal data with missing values in order to demonstrate how incorporating missing data mechanism uncertainty can increase the coverage of parameter estimates.

4.1. *Setup.* Building on an example in Hedeker and Gibbons [(2006), page 283], longitudinal data with missing values were simulated according to the following pattern-mixture model:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 \text{Time}_j + \beta_2 \text{Tx}_i + \beta_3(\text{Tx}_i \times \text{Time}_j) \\
&\quad + \beta_4(\text{Drop}_i \times \text{Time}_j) + v_{0i} + v_{1i}\text{Time}_j + \varepsilon_{ij},
\end{aligned} \tag{4.1}
$$

where Time_j was coded 0, 1, 2, 3, 4 for five timepoints, Tx_i was a dummy-coded (i.e., 0 or 1) grouping variable with 150 subjects in each group, and Drop_i was a dummy-coded variable indicating those subjects who eventually dropped out of the study. There were 100 dropouts in each treatment group. The regression coefficients were defined to be as follows: $\beta_0 = 25$, $\beta_1 = -3$, $\beta_2 = 0$, $\beta_3 = -1$, and $\beta_4 = 1.5$. This setup represents a randomized controlled trial in which group means are equal at baseline and there is a greater decrease in the outcome measure over time in the treatment group. Participants who eventually drop out of the study have smaller decreases in outcomes over time as compared to nondropouts. Thus, the slope of the treatment and control groups were -3.0 and -2.0, respectively. The random subject effects v_{0i} and v_{1i} were assumed normal with zero means, variances $\sigma^2_{v0} = 4$ and $\sigma^2_{v1} = 1$ and covariance $\sigma_{v01} = -0.1$. The errors ε_{ij} were assumed to be normal with mean 0 and variance $\sigma^2 = 9$ for nondropouts and $\sigma^2 = 16$ for dropouts.

We generated nonignorable missing values on y_{ij} using the following rule: at timepoints 1, 2, 3 and 4, subjects in the dropout group dropped out with probabilities (0.25, 0.50, 0.75, 1) so that the overall proportions of missing values were 0.17, 0.42, 0.60 and 0.67 for the four timepoints. Nondropouts have no missing values at any time point. The high proportion of dropouts and the use of monotone missingness (versus intermittent missingness) were chosen so that post-imputation inferences were sensitive to assumptions regarding the missing data mechanism.

Imputation using the multiplier approach of Section 3 proceeded as follows. We first generated 200 imputations of each missing value using the software package MICE [van Buuren and Oudshoorn (2011)] which imputes variables one-at-a-time based on a conditional distribution for each variable. We specified a linear regression model [Rubin (1987), page 166] which assumes the missing data are MAR. Each treatment group was imputed separately to preserve the desirable property in an intent-to-treat analysis framework that imputed values depend only on information from other cases in the same treatment arm.

Using the methods described in Sections 3, we then transformed the MICE imputations—which assume the data are ignorably missing—into imputations that assume the data are nonginorably misssing. Specifically, we simulated 100 values of k from one of the imputation model distributions listed

in Table 1 and described in Sections 4.2 and 4.3. Using equation (3.2), each of these values of k was multiplied to the imputed values in 2 imputed data sets to create 2 imputations nested within 100 models, that is, 200 imputed data sets.

We used $M = 100$ imputation models and $N = 2$ imputations within each model so that the degrees of freedom for the within-model variance $M(N-1)$ [equation (3.8)] and the degrees of freedom for the between-model variance $M-1$ [equation (3.9)] were approximately equal. This allowed us to estimate within- and between-model variance with equal precision, which is necessary for stable measurements of the rates of missing information [Harel (2007)].

We then analyzed the 200 imputed data sets using the random intercept and slope model described in equation (4.1) but without the covariates that include dropout. Inferences were combined using the nested multiple imputation combining rules described in Section 3.3. Here, for brevity, we focus on the slope of the treatment group.

One thousand replications for the above scenario were simulated. An R function for combining nested multiple imputation inferences and calculating rates of missing information is available in the supplementary materials [Siddique, Harel and Crespi (2012)].

4.2. *Ignorability assumptions.* We explored the effect of imputing under four different ignorability assumptions which we refer to as MAR, Weak NMAR, Strong NMAR and Misspecified NMAR. We now discuss each of these assumptions in turn:

(1) Missing at Random (MAR): Under this assumption, we generate multiple imputations assuming the data are missing at random. Specifically, we generate imputations assuming the multiplier k in equation (3.2) is drawn from a distribution with a mean of 1.0.

(2) Weak Not Missing at Random (Weak NMAR): Under this assumption, we generate multiple imputations assuming the data are not missing at random, but that nonrespondents are not very different from respondents. Specifically, imputations assuming weak NMAR are generated by assuming the multiplier k in equation (3.2) is drawn from a distribution with a mean of 1.3 (nonrespondents have values that are 30% larger than respondents).

(3) Strong NMAR: Here we generate multiple imputations assuming the data are NMAR and that nonrespondents are quite a bit different than respondents. Imputations are generated assuming nonrespondents are 70% larger than respondents (a multiplier distribution mean of 1.7).

(4) Misspecified NMAR: Here we generate multiple imputations assuming the data are NMAR but that nonrespondents have *lower* values than respondents even though in truth the reverse is true. Imputations assuming misspecified NMAR are generated by assuming the multiplier k in equation (3.2) is drawn from a distribution with a mean of 0.8 (nonrespondents have

values that are 20% smaller than respondents). We chose this assumption to demonstrate that even when the imputer is wrong about the nature of nonignorability, incorporating mechanism uncertainty can increase coverage and make a bad situation better.

4.3. *Mechanism uncertainty assumptions.* In addition to generating imputations using the above ignorability assumptions, we also generated imputations based on four different assumptions regarding how certain we were about the correctness of our models. When there is no mechanism uncertainty, all imputations are generated from the same model. When there is mechanism uncertainty, then multiple models are used. All models are centered around one of the ignorability assumptions in Section 4.2. Uncertainty is then characterized by departures from the central model. The four different uncertainty assumptions used to generate multiple models were as follows: no uncertainty, mild uncertainty, moderate uncertainty and ample uncertainty. These assumptions are described below:

(1) No uncertainty: This is the assumption of most imputation schemes. One imputation model is chosen and all imputations are generated from that one model. In particular, the most common imputation approach is to assume the data are MAR with no uncertainty. Imputations with no mechanism uncertainty were generated by using the same multiplier k in equation (3.2) for all 100 imputation models.

(2) Mild uncertainty: Here we assume that there is a small degree of uncertainty regarding what is the right mechanism. By incorporating some uncertainty into our choice of imputation model, imputations are generated using multiple models. Specifically, the multiplier k in equation (3.2) was drawn from a Normal distribution with a standard deviation of 0.1.

(3) Moderate uncertainty: Multiple models with moderate uncertainty are generated using equation (3.2) by drawing the multiplier from a Normal distribution with a standard deviation of 0.3.

(4) Ample uncertainty: Multiple models with ample uncertainty are generated using equation (3.2) by drawing the multiplier from a Normal distribution with a standard deviation of 0.5.

With four ignorability assumptions and four uncertainty assumptions, we imputed the data under a total of 16 scenarios. Within each scenario, we evaluated the percent bias and RMSE of the post-multiple-imputation treatment slope as well as the coverage rate and width of its nominal 95% interval estimate. In addition, we calculated measures of missing information: the overall estimated rate of missing information [$\hat{\gamma}$ in equation (3.13)], the estimated rate of missing information due to nonresponse [$\hat{\gamma}^w$ in equation (3.14)], the estimated rate of missing information due to model uncertainty, $\hat{\gamma}^b = \hat{\gamma} - \hat{\gamma}^w$, and the estimated contribution of model uncertainty to the overall rate of missing information as measured by the ratio $\frac{\hat{\gamma}^b}{\hat{\gamma}}$.

TABLE 1
Simulation study of multiple imputation of continuous data using multiple models. One hundred models, 2 imputations within each model

Ignore assump.	Uncertainty	Model Dist'n	PB	RMSE	Cvg.	Width of CI	$\hat{\gamma}$	$\hat{\gamma}^w$	$\hat{\gamma}^b$	$\frac{\hat{\gamma}^b}{\hat{\gamma}}$
MAR	None	$N(1.0, 0.0)$	33.04	1.01	0.1	0.75	0.63	0.62	0.01	0.02
	Mild	$N(1.0, 0.1)$	33.18	1.01	0.3	0.98	0.77	0.61	0.16	0.21
	Moderate	$N(1.0, 0.3)$	33.44	1.02	53.4	2.05	0.93	0.57	0.36	0.39
	Ample	$N(1.0, 0.5)$	33.72	1.03	99.5	3.28	0.96	0.49	0.47	0.49
Weak NMAR	None	$N(1.3, 0.0)$	18.22	0.59	36.2	0.96	0.64	0.63	0.01	0.02
	Mild	$N(1.3, 0.1)$	18.35	0.59	53.5	1.14	0.74	0.62	0.12	0.16
	Moderate	$N(1.3, 0.3)$	18.56	0.60	98.0	2.13	0.91	0.59	0.32	0.35
	Ample	$N(1.3, 0.5)$	18.77	0.61	100.0	3.33	0.95	0.53	0.42	0.44
Strong NMAR	None	$N(1.7, 0.0)$	−1.53	0.27	98.2	1.28	0.60	0.59	0.01	0.02
	Mild	$N(1.7, 0.1)$	−1.40	0.27	99.6	1.42	0.67	0.58	0.09	0.13
	Moderate	$N(1.7, 0.3)$	−1.19	0.27	100.0	2.29	0.86	0.56	0.29	0.34
	Ample	$N(1.7, 0.5)$	−1.03	0.28	100.0	3.42	0.92	0.53	0.40	0.43
Misspec. NMAR	None	$N(0.8, 0.0)$	42.95	1.30	0.0	0.64	0.57	0.56	0.01	0.02
	Mild	$N(0.8, 0.1)$	43.10	1.30	0.0	0.90	0.77	0.56	0.22	0.28
	Moderate	$N(0.8, 0.3)$	43.39	1.31	8.5	2.01	0.94	0.50	0.43	0.46
	Ample	$N(0.8, 0.5)$	43.70	1.32	88.1	3.26	0.96	0.43	0.54	0.56

PB: percent bias; RMSE: root mean squared error; Cvg: coverage.

5. Simulation results. Table 1 lists the results of our imputations under the 16 different ignorability/uncertainty scenarios using regression imputation and the methods described in Section 3 for the slope of the treatment group. Beginning with the first row, we see that assuming MAR with no mechanism uncertainty results in estimates that are highly biased with a coverage rate close to 0%. This result is not surprising, as the data are non-ignorably missing and here we are assuming in all of our models that the data are ignorably missing. Since we are using the same model for all imputations, $\hat{\gamma}^b$, the estimated fraction of missing information due to model uncertainty is approximately equal to 0 as is $\frac{\hat{\gamma}^b}{\hat{\gamma}}$, the estimated contribution of model uncertainty to the overall rate of missing information.

Moving to the subsequent rows in Table 1, still assuming MAR, we see the effect of increasing mechanism uncertainty on post-imputation parameter estimates. Both percent bias and RMSE are the same as with no uncertainty, but now coverage is increasing as we increase the amount of uncertainty in our imputation models. Coverage increases from 0% to 99.5%. The mechanism here is clear—by increasing the amount of uncertainty in our imputation models, we are now generating imputations under a range of ignorability assumptions. This additional variability in the imputed values translates to wider confidence intervals and hence greater coverage. We also see that our measures of missing information are able to pick up this uncertainty. Both

$\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ increase as the amount of model uncertainty increases. As model uncertainty increases, it becomes a larger proportion of the overall rate of missing information.

Since missing values in our simulation study tended to be larger than observed values, the weak and strong NMAR conditions result in smaller bias than the imputations assuming MAR. As before, increasing the amount of model uncertainty does not change bias but instead increases coverage (by increasing the width of the 95% confidence intervals) to the point that weak NMAR with moderate and ample uncertainty exceeds the nominal level. Under the strong NMAR assumption, bias is small enough that there is no benefit to additional mechanism uncertainty. Also, as before, additional model uncertainty is reflected in increasing values of $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$.

Finally, the last four rows of Table 1 present results when the missing data mechanism is misspecified. Here, the missing data are imputed assuming that missing values are smaller than observed values (even after conditioning on observed information) when in fact the reverse is true. Not surprisingly, bias and RMSE are poor in this situation, but by incorporating mechanism uncertainly into our imputations we are able to build some robustness into our imputation model. With ample uncertainly, coverage is 88.1%, a substantial increase over the coverage rate of 0%, which is the result of using the same (misspecified) model for all imputations.

6. Application to the Women Entering Care study. We applied our methods to the WECare data as follows. We imputed the continuous WECare HDRS scores using the same method and imputation model distribution parameters as described in the simulation study.

The Weak NMAR and Strong NMAR assumptions assume that missing values tend to be larger than observed values with the same covariates. Since higher HDRS scores reflect more depression symptoms, these assumptions imply that nonrespondents are more depressed than respondents even after conditioning on observed information. The term "Misspecified" NMAR is a misnomer in this setting because we do not actually know the correct specification. We use the term only to be consistent with the simulation study. For Misspecified NMAR, the assumption is that nonrespondents are less depressed than respondents.

We investigated how different factors in our imputation procedure affected inferences from the WECare data. In every scenario, 100 models were used and 2 imputations were generated within each model for every missing value. As in the simulation study, each treatment group was imputed separately.

When imputing and analyzing the WECare data, we restricted our attention to the depression outcomes that were analyzed in Miranda et al. (2003), variables used as covariates in final analyses, and a set of additional variables

TABLE 2
WECare variables used for imputation and analysis

Variable name	Imputation or analysis?	Percent missing	Variable type
Baseline HDRS	Both	0%	Scaled
Month 1 HDRS	Both	25%	Scaled
Month 2 HDRS	Both	24%	Scaled
Month 3 HDRS	Both	30%	Scaled
Month 4 HDRS	Both	34%	Scaled
Month 5 HDRS	Both	38%	Scaled
Month 6 HDRS	Both	30%	Scaled
Month 8 HDRS	Imputation	33%	Scaled
Month 10 HDRS	Imputation	34%	Scaled
Month 12 HDRS	Imputation	24%	Scaled
Ethnicity	Both	0%	Nominal
Age	Imputation	0%	Continuous
Income	Imputation	4%	Continuous
HS graduate	Imputation	0%	Binary
Number of children	Imputation	0%	Continuous
Received 9 wks of Meds	Imputation	0%	Binary (Med tx only)
No. of CBT sessions	Imputation	0%	Continuous (CBT tx only)
No. of mental health visits	Imputation	0%	Continuous (TAU tx only)
Insurance Status	Imputation	0%	Binary
Marital Status	Imputation	0%	Binary

HDRS: Hamilton depression rating scale.

used in the imputation models because they were judged to be potentially associated with the analysis variables. Table 2 lists variables that were used in imputation and analysis models and also indicates the percentage of missing values.

Four important targets of inference from the random intercept and slope model used in Miranda et al. (2003) are the slopes of the Medication treatment group and the CBT treatment group, reflecting the change in HDRS scores over time for the two active interventions and their difference with the slope of the TAU condition, which estimates the effect of treatment. Here, for brevity, we focus our attention on the slope of the Medication treatment group and also its difference with the slope of the TAU group (i.e., the Medication treatment effect) to illustrate the impact of different ignorability and uncertainty assumptions in our imputation procedures.

6.1. *Imputation of HDRS scores.* Imputation of the monthly HDRS scores using the multiplier approach of Section 3 proceeded as follows. For every ignorabilty/uncertainty combination in Table 1, we first generated 200 imputations of the WECare missing data using MICE [van Buuren and

Oudshoorn (2011)] and specified a linear regression model [Rubin (1987), page 166] to impute income and depression scores. This method assumes the missing data are MAR. Each imputation model conditioned on all the variables listed in Table 2. In particular, depression scores were imputed using a model that conditioned on both prior depression scores and subsequent depression scores in order to make use of all available information. Imputed values were rounded to the nearest observed value to create plausible HDRS scores.

We then simulated 100 values from the corresponding ignorability/uncertainty distributions listed in Table 1 and described in Sections 4.2 and 4.3. Using equation (3.2), each of these values of k was multiplied to the imputed values in 2 imputed data sets to create 2 imputations nested within 100 models. Many of the ignorability/uncertainty distributions that are used in the simulation are not realistic for this application, but we use them here for the sake of brevity and so that we can clearly see the effect of different assumptions on post-imputation inferences. Imputed values were again rounded to the nearest observed value to create plausible HDRS scores. We then analyzed the 200 imputed data sets using the random intercept and slope regression model of Miranda et al. (2003), and the nested imputation combining rules described in Section 3.4.

6.2. *Post multiple imputation results from the WECare analysis.* Table 3 provides estimates, standard errors, confidence intervals, p-values and rates of missing information for the WECare Medication slope by the 16 different ignorability/uncertainty scenarios described in Sections 4.2 and 4.3 using the multiple model approach described in Section 3. Table 4 provides the same information for the difference between the Medication and TAU slopes.

Looking first at Table 3, we see that assumptions regarding ignorability and uncertainty have an impact on parameter estimates and their associated standard errors. Starting with those rows assuming MAR, we see that the point estimate for the slope changes very little for all four uncertainty assumptions. However, as we assume more uncertainty, the associated standard errors increase. This same phenomenon was seen in the simulation study. The additional model uncertainty is also reflected in increasing values of $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$, the estimated rate of missing information due to model uncertainty and the estimated contribution of model uncertainty to the overall rate of missing information, respectively. These values are quite large under ample uncertainty, reflecting the fact that the ample uncertainty assumption is relatively diffuse for these data. Because of this, for every ignorability scenario, ample uncertainty results in slopes that are no longer significantly different from 0 at the 0.05 level.

As mentioned above, the Weak NMAR and Strong NMAR assumptions assume that nonrespondents are more depressed than respondents even after

TABLE 3

Post-imputation WECare Medication intervention slopes by ignorability/uncertainty scenario. One-hundred models with 2 imputations per model were used to generate 200 imputations. Multipliers were generated by drawing from a Normal distribution. MAR, Weak NMAR, Strong NMAR and Misspecified NMAR correspond to Normal distributions with means of 1, 1.3, 1.7 and 0.8, respectively. Amounts of uncertainty None, Mild, Moderate, Ample correspond to Normal distributions with standard deviations of 0, 0.1, 0.3 and 0.5, respectively

Ignore assump.	**Uncertainty**	**Est.**	**SE**	**LCI**	**UCI**	***p*-val.**	$\hat{\gamma}$	$\hat{\gamma}^w$	$\hat{\gamma}^b$	$\frac{\hat{\gamma}^b}{\hat{\gamma}}$
MAR	None	−1.93	0.47	−2.86	−1.00	<0.01	0.37	0.36	0.01	0.03
	Mild	−1.95	0.53	−3.00	−0.91	<0.01	0.49	0.37	0.13	0.25
	Moderate	−2.02	0.85	−3.70	−0.35	0.02	0.77	0.35	0.42	0.54
	Ample	−2.09	1.20	−4.46	0.28	0.08	0.87	0.32	0.54	0.63
Weak	None	−1.71	0.56	−2.81	−0.61	<0.01	0.42	0.41	0.01	0.03
NMAR	Mild	−1.74	0.60	−2.91	−0.57	<0.01	0.49	0.41	0.08	0.16
	Moderate	−1.82	0.84	−3.46	−0.17	0.03	0.72	0.40	0.33	0.45
	Ample	−1.91	1.15	−4.17	0.35	0.10	0.84	0.36	0.47	0.57
Strong	None	−1.53	0.65	−2.80	−0.25	0.02	0.42	0.40	0.01	0.03
NMAR	Mild	−1.54	0.66	−2.84	−0.24	0.02	0.45	0.40	0.04	0.09
	Moderate	−1.61	0.80	−3.19	−0.03	0.05	0.62	0.40	0.22	0.35
	Ample	−1.70	1.04	−3.74	0.34	0.10	0.76	0.38	0.39	0.51
Misspec.	None	−2.10	0.42	−2.93	−1.27	<0.01	0.30	0.29	0.01	0.03
NMAR	Mild	−2.12	0.49	−3.09	−1.16	<0.01	0.47	0.30	0.17	0.37
	Moderate	−2.18	0.85	−3.85	−0.51	0.01	0.79	0.29	0.49	0.63
	Ample	−2.22	1.20	−4.59	0.16	0.07	0.87	0.28	0.59	0.68

SE: standard error; LCI: lower 95% confidence interval; UCI: upper 95% confidence interval.

conditioning on observed information. Since there are more missing values later in the study, these assumptions have the effect of flattening the slope of the Medication intervention. Within any ignorability assumption, the point estimates of the slope change only a little but standard errors increase as more model uncertainty is assumed. Again, the values of $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ appear to capture this uncertainty.

The "Misspecified" NMAR assumption assumes that nonrespondents are less depressed than respondents and, as a result, the slope estimate is steeper than any of the other scenarios.

Table 4 displays results for the difference between the Medication and TAU slopes. For this quantity, the point estimate is almost the same in every ignorability/uncertainty scenario. This result is not surprising, as there were similar amounts of missing Medication and TAU data at each timepoint. For each ignorability assumption, the slope of the TAU intervention changed by the same magnitude as the slope of the Medication intervention.

TABLE 4

Post-imputation WECare Medication intervention treatment effects by ignorability/uncertainty scenario. One hundred models with 2 imputations per model were used to generate 200 imputations. Multipliers were generated by drawing from a Normal distribution. MAR, Weak NMAR, Strong NMAR and Misspecified NMAR correspond to Normal distributions with means of 1, 1.3, 1.7 and 0.8, respectively. Amounts of uncertainty None, Mild, Moderate, Ample correspond to Normal distributions with standard deviations of 0, 0.1, 0.3 and 0.5, respectively

Ignore assump.	**Uncertainty**	**Est.**	**SE**	**LCI**	**UCI**	***p*-val.**	$\hat{\gamma}$	$\hat{\gamma}^w$	$\hat{\gamma}^b$	$\frac{\hat{\gamma}^b}{\hat{\gamma}}$
MAR	None	−0.69	0.25	−1.18	−0.19	<0.01	0.34	0.34	0.00	0.00
	Mild	−0.69	0.27	−1.22	−0.17	0.01	0.42	0.35	0.06	0.15
	Moderate	−0.70	0.38	−1.46	0.05	0.07	0.69	0.34	0.35	0.51
	Ample	−0.71	0.52	−1.73	0.31	0.17	0.81	0.31	0.49	0.61
Weak	None	−0.70	0.30	−1.29	−0.11	0.02	0.37	0.37	0.00	0.00
NMAR	Mild	−0.71	0.31	−1.31	−0.10	0.02	0.41	0.38	0.02	0.05
	Moderate	−0.71	0.39	−1.48	0.05	0.07	0.62	0.37	0.25	0.40
	Ample	−0.72	0.51	−1.72	0.29	0.16	0.77	0.35	0.41	0.54
Strong	None	−0.70	0.35	−1.39	−0.00	0.05	0.36	0.36	0.00	0.00
NMAR	Mild	−0.70	0.35	−1.39	−0.01	0.05	0.37	0.37	0.00	0.00
	Moderate	−0.71	0.40	−1.49	0.07	0.07	0.50	0.38	0.12	0.24
	Ample	−0.71	0.48	−1.66	0.23	0.14	0.66	0.37	0.29	0.44
Misspec.	None	−0.67	0.22	−1.12	−0.23	<0.01	0.27	0.27	0.00	0.00
NMAR	Mild	−0.68	0.25	−1.16	−0.20	<0.01	0.38	0.29	0.10	0.26
	Moderate	−0.69	0.38	−1.43	0.05	0.07	0.71	0.29	0.42	0.59
	Ample	−0.70	0.52	−1.72	0.32	0.18	0.82	0.27	0.55	0.67

SE: standard error; LCI: lower 95% confidence interval; UCI: upper 95% confidence interval.

As a result, their difference remains constant at each assumption. However, incorporating model uncertainty into the imputations does increase the standard error of this parameter estimate. In fact, under moderate and ample uncertainty the treatment effect of the Medication intervention is no longer significant at the 0.05 level. These results underscore the importance of making reasonable assumptions. As noted above, the uncertainty assumptions in this example were chosen to be consistent with the simulation study and may not be realistic in a depression study.

In the scenarios in Table 4 where there was no model uncertainty, the original estimates of the rate of missing information due to model uncertainty were negative. As noted by Harel and Stratton (2009), this is possible due to the use of the method of moments for calculating the rates of missing information. Following their recommendation, we set $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$ equal to 0 when $\hat{\gamma}^b$ was negative.

7. Discussion. We have described a relatively simple method for generating multiple imputations in the presence of nonignorable nonresponse. By generating multiple imputations from multiple models, our method allows the user to incorporate uncertainty regarding the missing data mechanism into their parameter estimates. This is a useful approach when the missing data mechanism is unknown, which is almost always the case with nonignorably missing data. Our goal was not to develop a competitor to model-based methods such as selection models and pattern-mixture models. Instead, we wished to provide a imputation-based alternative to model-based methods for those researchers who prefer to use complete-data methods.

As seen in both the simulation studies and the application to the WECare data, post-imputation inferences can be highly sensitive to the choice of the imputation model. With the WECare data, imputation using our methods had a strong effect on the slope of the Medication intervention but little effect on the difference in slopes between the Medication and TAU groups. However, the Medication treatment effect was no longer significant when moderate and ample imputation model uncertainty were assumed.

This ability to render nonsignificant a result that is significant assuming ignorability (and vice versa) suggests that careful attention should be paid to the specification of the imputation model in equation (3.3). It may make sense to have analysis protocols specify clearly in advance what missing data assumptions will be explored. Imputation model assumptions should be chosen prior to analysis and not based on whether it produces the desired result. Here, the literature on prior elicitation may be helpful [Kadane and Wolfson (1998), Paddock and Ebener (2009), White et al. (2007)].

One approach for eliciting expert opinion when choosing a distribution for the multiplier k in equation (3.2) is to ask a subject-matter expert to provide an upper and lower bound for the multiplier. Then, assuming the multiplier is normally distributed, set the multiplier distribution mean equal to the average of the lower and upper bounds, and the standard deviation equal to the difference in bounds divided by 4. This assumes that the range defined by the upper and lower bounds is a 95% confidence interval which may be appropriate given the tendency of people to specify overly narrow confidence intervals [Tversky and Kahneman (1974)]. A similar calculation can be used if assuming a uniform prior.

Once the data have been imputed, it is important to examine rates of missing information, in particular, $\hat{\gamma}^b$ and $\frac{\hat{\gamma}^b}{\hat{\gamma}}$, to confirm that appropriate uncertainty is being incorporated into imputations. For example, if imputations outside the range of possible values are rounded up or down to the nearest observed value, this could result in too little variability, resulting in decreased coverage.

One approach for ensuring that appropriate uncertainty is incorporated into inferences is to generate imputations and perform analyses based on a

few different distributions for the multiplier. Then, without examining the significance of parameter estimates, confirm that appropriate imputation model uncertainty is being incorporated into the parameter estimates. Because our methods begin with the same set of ignorable imputations, it is relatively easy to generate imputations using different missing data mechanisms.

Our approach uses a large number of imputation models M, as this is necessary to obtain stable estimates of the rates of missing information. The relative (compared to an infinite number of imputations) efficiency of point estimates using nested multiple imputation is a function of the fraction of missing information as well as M and N. Improvements in relative efficiency are minimal when one uses more than a modest number of imputations. Hence, when the researcher's main interest is point estimates and their variances, a smaller number of imputations are usually sufficient, for example, $M = 10$–20 and $N = 2$ [Harel (2007)].

In line with more of a sensitivity analysis rather than a final analysis, when it is hard to pin down a single range for the multiplier, one may consider a growing set of ranges and observe how subsequent inferences evolve accordingly. This approach will allow the user to make more precise statements regarding the exact conditions under which the obtained results apply [van Buuren, Boshuizen and Knook (1999)].

Although we believe that all imputation model uncertainty should be incorporated into one inference, our approach is not inconsistent with a sensitivity analysis that examines inferences across a range of ignorability assumptions. Scharfstein, Rotnitzky and Robins (1999) view sensitivity analysis as useful "preprocessing" for any full Bayesian analysis that places prior distributions on sensitivity parameters and recommend that one also publish the results based on the individual sensitivity parameters in addition to the results that average across a range of sensitivity parameters so that readers are aware of how inferences vary based on individual sensitivity parameters.

Our approach is less extreme than worst-case best-case intervals [Cochran (1977), page 361] because we allow for imputation model parameters to fall within a chosen range in order to obtain narrower and more plausible ranges of estimates. Including implausible imputation model parameters broadens the range of inferences unnecessarily and can introduce implausible values. Instead, our imputation models are given appropriate weight, with imputation models that lead to extreme scenarios receiving less weight than models that lead to less extreme alternatives.

Of course, in any applied setting it is impossible to know exactly how strong a nonignorable assumption one should make and how much uncertainty one should place on their models. We see the second of these dilemmas—incorporating appropriate mechanism uncertainty—as deserving more attention. Attempting to correctly specify the missing data mechanism

is difficult in most settings. Still, we see our method as an improvement over methods that make no assumptions regarding missing data mechanism uncertainty. In addition, our method provides easily stated subjective notions regarding nonresponse so that they can be easily stated, communicated and compared.

We see a number of possible variations of our approach. For example, in some longitudinal data settings, it may be appropriate to use ignorable models early in the study, and nonignorable models later in the study, or perhaps incorporate less mechanism uncertainty early in the study and more later in the study.

Another possible approach is to use different imputation models for different groups of participants. For example, in the WECare study, we might want to generate nonignorable imputations for dropouts and ignorable imputations for everyone else. If the reasons for missingness are thought to differ by treatment group, it may be appropriate to use different assumptions for each treatment group. If one believes that nonresponse is due to both NMAR and MAR mechanisms [Barnes et al. (2010)], one could draw the multiplier from a mixture of distributions centered around both MAR and NMAR assumptions.

When an analyst has prior beliefs about the nature of missingness at a given time point given what occurred at previous time points, careful thought should go into the choice of the imputation model and multiplier distribution. Uncertainty regarding these beliefs can also be incorporated into the multiple models framework. Alternatively, methods that explicitly model this temporal relationship such as selection models and pattern-mixture models may be more appropriate [Molenberghs et al. (2003), Thijs et al. (2002)].

Some other approaches for generating multiple-model multiple imputations that can be incorporated into our framework include mixture model imputation [Rubin (1987), van Buuren, Boshuizen and Knook (1999)], imputation based on a multivariate t-distribution with varying degrees of freedom [Liu (1995)] and pattern-mixture model imputation [Demirtas and Schafer (2003), Thijs et al. (2002)]. Carpenter, Kenward and White (2007) propose an extension to their method where the multiple reweighting parameters are drawn from a Normal distribution to incorporate uncertainty in the sensitivity parameter. Finally, a nonignorable approximate Bayesian bootstrap [Rubin and Schenker (1991), Siddique and Belin (2008b)] in conjunction with hot-deck imputation can be also be used. This approach has the added benefit of generating plausible imputed values since imputations are based on values observed elsewhere. An important consideration when developing methods for generating nonignorable imputations is that as the methods become more complex, it becomes harder to communicate exactly how imputations were generated and the payoff for the additional complexity is not always clear.

APPENDIX: MOTIVATION FOR USING NESTED MULTIPLE IMPUTATION

In this section we provide motivation for using the nested multiple imputation combining rules. As in Section 3, let Q be the quantity of interest, Y_{mis} represent the missing values and ψ the imputation model. The observed data posterior of Q using our approach is

$$\begin{aligned} p(Q|Y_{\text{obs}}) &= \int\int p(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)p(Y_{\text{mis}}, \psi|Y_{\text{obs}})\,dY_{\text{mis}}\,d\psi \\ &= \int\int p(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)p(Y_{\text{mis}}|Y_{\text{obs}}, \psi)p(\psi)\,dY_{\text{mis}}\,d\psi. \end{aligned} \tag{A.1}$$

Note the posterior distribution of Y_{mis}, $p(Y_{\text{mis}}|\psi, Y_{\text{obs}})$, conditions on ψ so that nested multiple imputations are not independent draws from the same posterior distribution. When the posterior mean and variance are adequate summaries of the posterior distribution, equation (A.1) can be effectively replaced by

$$E(Q|Y_{\text{obs}}) = E(E(E(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)|Y_{\text{obs}}, \psi)) \tag{A.2}$$

and

$$\begin{aligned} \text{Var}(Q|Y_{\text{obs}}) &= E(\text{Var}(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)) + \text{Var}(E(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)) \\ &= E(E(\text{Var}(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)|Y_{\text{obs}}, \psi)) \qquad \text{(A.3)} \\ &\quad + E(\text{Var}(E(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)|Y_{\text{obs}}, \psi)) \qquad \text{(A.4)} \\ &\quad + \text{Var}(E(E(Q|Y_{\text{obs}}, Y_{\text{mis}}, \psi)|Y_{\text{obs}}, \psi)). \qquad \text{(A.5)} \end{aligned}$$

The three variance components in equations (A.3), (A.4) and (A.5) correspond to the the overall average complete data variance, the within-model variance and the between-model variance, respectively.

The mean in equation (A.2) is approximated using equation (3.5). And the variance components in equations (A.3), (A.4) and (A.5) are approximated using equations (3.7), (3.8) and (3.9) in Section 3.4.

Acknowledgments. The authors wish to thank the Editor and two anonymous reviewers whose comments greatly improved the quality of the manuscript. The authors also wish to thank Jeanne Miranda for use of the WE-Care data.

SUPPLEMENTARY MATERIAL

CombineNestedImputations: An R function for combining inferences based on nested multiple imputations (DOI: 10.1214/12-AOAS555SUPP; .R). This R function combines inferences based on nested multiply imputed data sets and calculates rates of missing information.

REFERENCES

Barnes, S. A., Larsen, M. D., Schroeder, D., Hanson, A. and Decker, P. A. (2010). Missing data assumptions and methods in a smoking cessation study. *Addiction* **105** 431–437.

Belin, T. R., Diffendal, G. J., Mack, S., Rubin, D. B., Schafer, J. L. and Zaslavsky, A. M. (1993). Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation (with discussion). *J. Amer. Statist. Assoc.* **88** 1149–1166.

Blackburn, I. M., Bishop, S., Glen, A. I., Whalley, L. J. and Christie, J. E. (1981). The efficacy of cognitive therapy in depression: A treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *Br. J. Psychiatry* **139** 181–189.

Carpenter, J. R., Kenward, M. G. and White, I. R. (2007). Sensitivity analysis after multiple imputation under missing at random: A weighting approach. *Stat. Methods Med. Res.* **16** 259–275. MR2371009

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York–London–Sydney. MR0474575

Collins, L. M., Schafer, J. L. and Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* **6** 330–351.

Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies. Monographs on Statistics and Applied Probability* **109**. Chapman & Hall/CRC, Boca Raton, FL. MR2459796

Demirtas, H. and Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Stat. Med.* **22** 2553–2575.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J. and Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Arch. Gen. Psychiatry* **46** 971–982.

Forster, J. J. and Smith, P. W. F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 57–70. MR1625664

Harel, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Stat. Methodol.* **4** 75–89. MR2339010

Harel, O. (2008). Outfluence—the impact of missing values. *Model Assist. Stat. Appl.* **3** 161–168. MR2518797

Harel, O. (2009). *Strategies for Data Analysis with Two Types of Missing Values: From Theory to Application*. Lambert Academic Publishing, Saarbrücken, Germany.

Harel, O. and Stratton, J. (2009). Inferences on the outfluence—how do missing values impact your analysis? *Comm. Statist. Theory Methods* **38** 2884–2898. MR2568193

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Wiley-Interscience, Hoboken, NJ. MR2284230

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: A review. *TEST* **18** 1–43. MR2495958

Kaciroti, N. A., Raghunathan, T. E., Schork, M. A., Clark, N. M. and Gong, M. (2006). A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: Evaluation of an asthma education program. *J. Amer. Statist. Assoc.* **101** 435–446. MR2256165

Kadane, J. and Wolfson, L. J. (1998). Experiences in elicitation. *J. Roy. Statist. Soc. Ser. D* **47** 3–19.

Landrum, M. B. and Becker, M. P. (2001). A multiple imputation strategy for incomplete longitudinal data. *Stat. Med.* **20** 2741–2760.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley-Interscience, Hoboken, NJ. MR1925014

Liu, C. (1995). Missing data imputation using the multivariate t distribution. *J. Multivariate Anal.* **53** 139–158. MR1333132

Miranda, J., Chung, J. Y., Green, B. L., Krupnick, J., Siddique, J., Revicki, D. A. and Belin, T. (2003). Treating depression in predominantly low-income young minority women: A randomized controlled trial. *JAMA* **290** 57–65.

Molenberghs, G., Kenward, M. G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: The Slovenian plebiscite case. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 15–29.

Molenberghs, G., Thijs, H., Kenward, M. G. and Verbeke, G. (2003). Sensitivity analysis of continuous incomplete longitudinal outcomes. *Stat. Neerl.* **57** 112–135. MR2035862

Paddock, S. M. and Ebener, P. (2009). Subjective prior distributions for modeling longitudinal continuous outcomes with non-ignorable dropout. *Stat. Med.* **28** 659–678. MR2655736

Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V. and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* **27** 85–95.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J. Amer. Statist. Assoc.* **72** 538–543. MR0445668

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. MR0899519

Rubin, D. B. (2003). Nested multiple imputation of NMES via partially incompatible MCMC. *Stat. Neerl.* **57** 3–18. MR2055518

Rubin, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Stat. Med.* **10** 585–598.

Rubin, D. B., Stern, H. S. and Vehovar, V. (1995). Handling "don't know" survey responses: The case of the Slovenian plebiscite. *J. Amer. Statist. Assoc.* **90** 822–828.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data. Monographs on Statistics and Applied Probability* **72**. Chapman & Hall, London. MR1692799

Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychol. Methods* **7** 147–177.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478

Schenker, N. and Taylor, J. M. G. (1996). Partially parametric techniques for multiple imputation. *Comput. Statist. Data Anal.* **22** 425–446.

Shen, Z. J. (2000). Nested multiple imputation. Ph.D. thesis, Dept. Statistics, Harvard Univ., Cambridge, MA.

Siddique, J. and Belin, T. R. (2008a). Multiple imputation using an iterative hot-deck with distance-based donor selection. *Stat. Med.* **27** 83–102. MR2416864

Siddique, J. and Belin, T. R. (2008b). Using an approximate Bayesian bootstrap to multiply impute nonignorable missing data. *Comput. Statist. Data Anal.* **53** 405–415. MR2649095

SIDDIQUE, J., HAREL, O. and CRESPI, C. M. (2012). Supplement to "Addressing missing data mechanism uncertainty using multiple-model multiple imputation: Application to a longitudinal clinical trial." DOI:10.1214/12-AOAS555SUPP.

THIJS, H., MOLENBERGHS, G., MICHIELS, B., VERBEKE, G. and CURRAN, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3** 245–265.

TVERSKY, A. and KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* **185** 1124–1131.

VAN BUUREN, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **16** 219–242. MR2371007

VAN BUUREN, S., BOSHUIZEN, H. C. and KNOOK, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.* **18** 681–694.

VAN BUUREN, S. and OUDSHOORN, C. (2011). MICE: Multivariate Imputation by Chained Equations. R package version 2.5.

VANSTEELANDT, S., GOETGHEBEUR, E., KENWARD, M. G. and MOLENBERGHS, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sinica* **16** 953–979. MR2281311

WACHTER, K. W. (1993). Comment on hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation. *J. Amer. Statist. Assoc.* **88** 1161–1163.

WARDEN, D., RUSH, A. J., WISNIEWSKI, S. R., LESSER, I. M., KORNSTEIN, S. G., BALASUBRAMANI, G. K., THASE, M. E., PRESKORN, S. H., NIERENBERG, A. A., YOUNG, E. A., SHORES-WILSON, K. and TRIVEDI, M. H. (2009). What predicts attrition in second step medication treatments for depression?: A STAR*D report. *The International Journal of Neuropsychopharmacology* **12** 459–473.

WHITE, I. R., CARPENTER, J., EVANS, S. and SCHROTER, S. (2007). Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clinical Trials* **4** 125–139.

J. SIDDIQUE
DEPARTMENT OF PREVENTIVE MEDICINE
NORTHWESTERN UNIVERSITY
FEINBERG SCHOOL OF MEDICINE
CHICAGO, ILLINOIS 60611
USA
E-MAIL: siddique@northwestern.edu

O. HAREL
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269
USA
E-MAIL: oharel@stat.uconn.edu

C. M. CRESPI
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF CALIFORNIA LOS ANGELES
SCHOOL OF PUBLIC HEALTH
LOS ANGELES, CALIFORNIA 90095
USA
E-MAIL: ccrespi@ucla.edu